Building a GAN for Replicating Epithelial Impedance Spectra for ML-based Pattern Recognition

Lena Jurkschat, Benjamin Schindler

Abstract: Impedance spectroscopy is a common method in the field of biotechnology to measure electrical conductivity of special cell lines (i.e. ephitelial). Based on the measured impedance spectra, machine learning (ML) techniques including random forests and feedforward networks are increasingly used to determine physiological properties of the underlying cell tissue and to detect a wide range of diseases. However, training ML models for this purpose typically requires large amounts of data and real cell tissue measurements are costly to obtain due to their experimental setup. This paper introduces a Generative Adversarial Network (GAN) which meets the high demand for training data by replicating impedance spectra from a given data set. As a proof of concept, we show that GANs are capable of generating spectra that have a similar shape to the original ones and could therefore be used to overcome a lack of training data.

Keywords: GAN, impedance spectroscopy, neural networks, epithelia

1 Introduction

Artificial Neural Networks (ANNs) cover a variety of tasks, such as classification, regression and translation of texts. A special kind of ANNs are so called *Generative Adversarial Networks* (*GAN*), which are composed of a *Generator* and a *Discriminator* subnetwork. As generative models, GANs learn distributions within a data set and, if trained successfully, are able to generate new samples from them. GANs are very well known for their capacity of generating fake pictures of persons [Goo+14]. Trained on large sets of high-dimensional data (e.g. pixels of an input image), a GAN is able to generate fake pictures that a human cannot distinguish from real ones.

In the field of biotechnology, impedance spectroscopy is widely used in the investigation of epithelial tissues to determine their electrical conductivity and the associated flow of certain ions using so called Ussing Chambers [LSH04]. Furthermore, measured impedance spectra reveal information about physiological properties of the underlying cell tissue (e.g. membrane capacitance and subepithelial resistance) and can be used for various medical applications, such as detecting breast cancer [Rah+20] and muscular damage [Mły+19]. In the last decade, various cell cultures, e.g. HT29/B6 and IPEC-J2, were studied using impedance spectroscopy, in which the complex valued impedance is determined as a function of the frequency of an alternating current [Gün+12].

It has been shown that machine learning algorithms are able to analyse impedance spectra and approximate the sought physiological properties through patterns in the data [SBG13].

For this purpose, impedance spectra have been modeled to overcome the lack of expensive measured data, which are needed in a great amount for the training process [Sch18]. Thus, high accuracies in the determination could be achieved, but new synthesis methods are needed that allow a more realistic modeling of the cell tissues based on real measurements. Here we introduce a new approach to replicate impedance spectra using GANs in order to enlarge the data quantity. In this paper, synthetic data is used for this purpose, but the principle procedure could also be easily applied to real impedance measurements in the future.

This paper describes the implemented GAN and presents some of the generated spectra. Although the dimension of the data is comparatively low (84 values per sample against pictures with thousands of pixels), it was possible to create a GAN that accomplishes that challenge. We present an architectural small network that is able to create not distinguishable artificial impedance spectra from four randomly chosen values of latent space.

Section 2 introduces the data and in section 3, the functionality of the GAN is explained. The generated impedance spectra is presented and evaluated in section 4. A conclusion of the approach is given in section 5.

2 Data

For this work, a data set of modelled impedance measurements on the epithelial cell line HT29/B6 under physiological control conditions was adopted from [Sch20]. HT29/B6 is a well-studied carcinoma cell culture derived from the human colon. The synthesised data is based on experimentally estimated value ranges and additional error modelling which are both obtained from [Sch18]. The data set has the advantage that it has already been used to predict epithelial properties using machine learning and is extensively characterised (cf. [Sch20]).

The data set includes $150,000$ spectra, each consisting of 42 measurements taken at different frequencies ($1.3Hz - 16350Hz$). As complex values, impedances comprise real and imaginary parts. Impedance spectra are therefore often displayed in Nyquist representation, in which the imaginary part is plotted against the real part (Fig. 2). Looking at real and imaginary parts as separate features, this results in 84 values as input for the neural network.

3 Generative Adversarial Networks

Generative Adversarial Networks (GAN) are a special kind of Artificial Neural Networks [Goo+14]. Generally it is build out of two separate models, that are trained against each other.
There is a generator model G that takes random noise as input and is supposed to produce realistic but synthetic data as output, in our case the generated impedance spectra. The

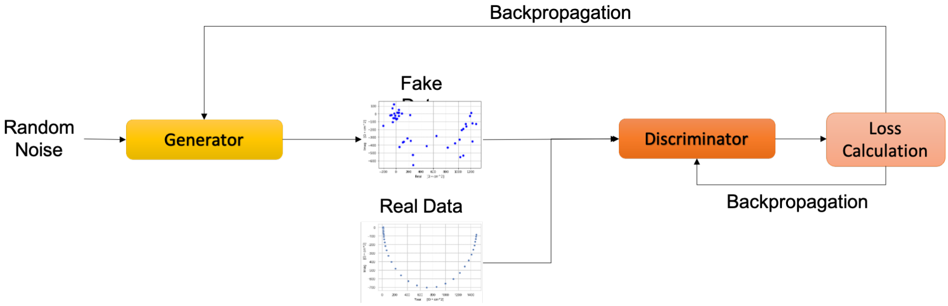


Fig. 1: Overview of GAN functionality, here for impedance spectra

discriminator model D then distinguishes whether the given data belongs to the original data or is fake data from the generator. The error of that model (i.e. incorrect identified fake and real data) is then propagated to itself and the generator to train its recognition of fake data and the production of synthetic data respectively. Here, the Binary Cross Entropy Loss (BCE Loss) was used as the error function, calculated on the basis of the correctly and incorrectly categorized data.
An overview of the GAN training for the special case of impedance spectra can be found in figure 1.

The model was build using the *PyTorch*[1] neural network framework. The generator model in this paper uses four randomly generated numbers (i.e. Noise) approximately in the range of the impedances as an input for the generator to generate the impedance spectra.

Compared to many state-of-the-art ANNs, its architecture is relatively small. It was build as a fully connected feedforward neural network, using a combination of Linear and LeakyReLU (Leaky Rectified Linear Unit) layers [NH10] as a base model for both, generator and discriminator. Additionally the discriminator uses the sigmoid function at the ouput layer for classification into real and fake. For testing different architectures, an ELU (Exponential Linear Unit) has been added to the generator of the base model. The differences of both models are discussed in section 4.
However, ultimately 5 Layers for the generator were used and 7 for the discriminator including input and output layer, so its training time was short although no GPU was used. The Generator uses $[4, 8, 16, 42, 84]$ neurons belonging to the layers and the Discriminator $[84, 64, 32, 16, 8, 4, 1]$ respectively. The complete model was trained with all given data samples and for 30 epochs with a batch size of 32, optimized using the AdaGrad optimizer [DHS11] with an initial learning rate of 0.001.

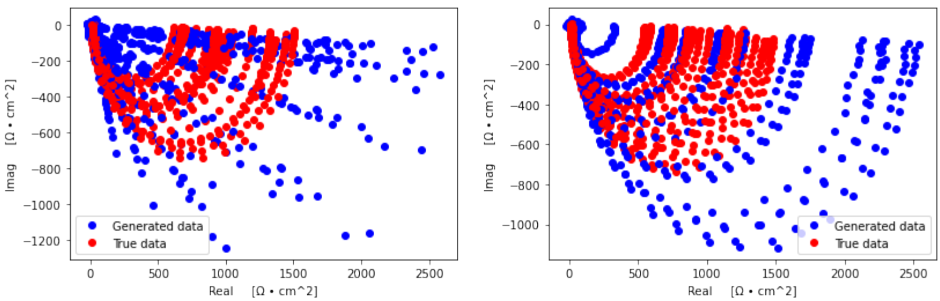
[1] `https://pytorch.org/docs/stable/nn.html`

4 Results

Within the process, two major GANs have been trained. Firstly, one GAN without an ELU Unit within the Generator and secondly, one where the ELU Unit has been added to the second-to-last layer due to the already successful training of other GANs [Agg+19]. For the target of generating new impedance spectra, the generator functionality has to be proofed. Therefore, the Results of the generator models have been compared graphically to the training data (Fig. 2). Twenty generated impedance spectra and another twenty randomly selected spectra from the original data set were chosen to be plotted.

While testing various units, the introduction of ELUs instead of solely using ReLUs was the key to train the right contexts from the data. This improvement is shown in Figure 2. Utilizing only ReLU units led to a generator model that learns a more linear structure than the typical semi-circle shape it is supposed to generate (Fig. 2(a)). Using an ELU Unit additionally, it turns out that the generated data has the supposed semi-circle shape but scales wider in its value range (Fig. 2(b)).

To validate the training process, the GAN losses were also recorded. Within the process, the losses of the generator and discriminator should adapt to each other as the generator improves and discriminator has a harder task to differentiate true from fake data. At the end of the process, the discriminator is not able to distinguish fake from real, so the discriminator gets ~ 50% right. As seen in figure 3, the Generator starts with a higher loss, but adapts to the discriminator, as the generation of fake data gets better. Obviously, the discriminator behaves vice versa. Note that it's typical for a GAN that the loss is rising in the beginning due to the adversarial training.



(a) First Approach using a LeakyReLU Unit (b) Second Approach including an ELU Unit

Fig. 2: Comparison of original (red) and replicated (blue) impedance spectra generated with a GAN using only LeakyReLU layers (a) vs. additional ELU layers (b).

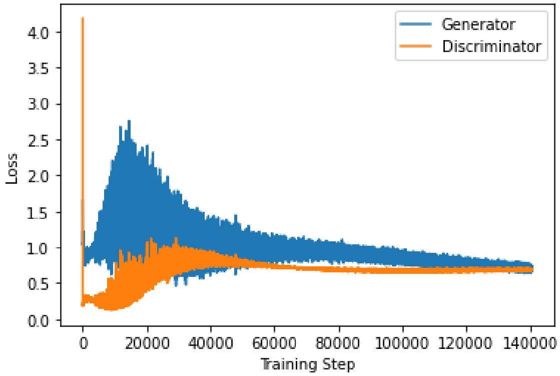


Fig. 3: Loss of the ELU Discriminator and Generator Model during the training process.

5 Conclusion

It has been shown that Generative Adversarial Networks are capable of replicating impedance spectra from modeled data. Even though the generated data scales wider then the true one (fig. 2(b)), it creates the right shape for various new data. Further data sets have to be investigated to validate, if this model is able to replicate spectra for other cell lines and cell conditions with more complex impedance curves.
Therefore, the method presented here can already be used to enlarge impedance data sets. However, the procedure developed in this paper must be seen as a first step. For future work, the won spectra have to be analyzed with proper quantitative metrics to determine if they are reliable compared to measured spectra. Additionally, the impedance curves must be mapped with the corresponding cell model parameters, in order to use the won data for regression tasks and pattern recognition with biological application such as in [SBG13].

Acknowledgement

We thank Thomas Schmid from the Machine Learning Group of Leipzig University for his supervision in the field of epithelial impedance spectroscopy, for providing the data set and for his beneficial advice during the implementation of the network.

References

[LSH04] Hongyu Li, David N Sheppard, and Martin J Hug. "Transepithelial electrical measurements with the Ussing chamber". In: *Journal of Cystic Fibrosis* 3 (2004), pp. 123–126.

[NH10] Vinod Nair and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, 2010, pp. 807–814. ISBN: 9781605589077.

[DHS11] John Duchi, Elad Hazan, and Yoram Singer. "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". In: *Journal of Machine Learning Research* 12.61 (2011), pp. 2121–2159. URL: `http://jmlr.org/papers/v12/duchi11a.html`.

[Gün+12] Dorothee Günzel et al. "From TER to trans- and paracellular resistance: lessons from impedance spectroscopy". In: *Annals of the New York Academy of Sciences* 1257 (June 2012), pp. 142–151. ISSN: 0077-8923. DOI: `10.1111/j.1749-6632.2012.06540.x`. URL: `https://doi.org/10.1111/j.1749-6632.2012.06540.x`.

[SBG13] Thomas Schmid, Martin Bogdan, and Dorothee Günzel. "Discerning Apical and Basolateral Properties of HT-29/B6 and IPEC-J2 Cell Layers by Impedance Spectroscopy, Mathematical Modeling and Machine Learning". In: *PLOS ONE* 8.7 (July 2013), pp. 1–12. DOI: `10.1371/journal.pone.0062913`. URL: `https://doi.org/10.1371/journal.pone.0062913`.

[Goo+14] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014. URL: `https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf`.

[Sch18] Thomas Schmid. "Automatisierte Analyse vonImpedanzspektren mittelskonstruktivistischen maschinellen Lernens". Dissertation. Universität Leipzig, 2018. URL: `https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-318388`.

[Agg+19] Karan Aggarwal et al. "Regression with Conditional GAN". In: *arXiv preprint arXiv:1905.12868* (May 2019).

[Mły+19] Marcel Młyńczak et al. "Impedance Spectroscopy Method to Detect Pelvic Floor Muscle Damage—A Feasibility Study". In: *World Congress on Medical Physics and Biomedical Engineering 2018*. Singapore: Springer Singapore, 2019, pp. 875–878. ISBN: 978-981-10-9038-7.

[Rah+20] Sam Matiur Rahman et al. "Ensemble-Based Machine Learning Algorithms for Classifying Breast Tissue Based on Electrical Impedance Spectroscopy". In: *Advances in Artificial Intelligence, Software and Systems Engineering*. Cham: Springer International Publishing, 2020, pp. 260–266. ISBN: 978-3-030-20454-9.

[Sch20] Benjamin Schindler. "Maschinelles Lernen zur paarweisen Analyse epithelialer Impedanzspektren". German. MA thesis. Universität Leipzig, 2020. URL: `https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-742028`.